ANALYSIS SUMMARY

By the summary statistics and boxplots mean, median, standard deviation, variance and outliers are observed.
With the help of correlation matrix and corrgram, hypotheses of mutual independence (Significance of Relationships) can be drawn.
These variables can be tested further via t test etc.
I have created airline$AIRLINE variable to numerical class, so that correlation matrix can be observed in full data set.

Correlation -

Flight duration - Price economy, Price premium
International - Pitch premium, width premium, diffPITCH
Seats economy- seats premium, N
Seats premium - seats economy, N
Pitch economy - -pitch premium, -width premium, -diffPITCH
Pitch premium - -pitch economy, international, width premium
Width economy- none
Width premium - airline, international, pitch premium,Relative
Price economy - flight duration, price premium
Price premium - price economy , Flight duration
N - sets economy, seats premium
LAMBDA- none

Regression model -

Now, ***what factors explain the difference in price between an economy ticket and a premium-economy airline ticket?***

For understanding factors relating in Difference between economy ticket and premium economy ticket price. First we need to understand Price of Premium economy ticket.

We have economy price of ticket now premium economy depends ticket price will depend on economy ticket price ,because it's modification of that class. Available factors in our data set which shows difference between two classes are width, pitch , no of seats.
And it is generally noticed that mostly premium economy are available in international flights which most probably have big flight duration.
So Flight duration and International are also variables to consider for

regression for Price premium.
We have to understand difference between two classes , so actual width and pitches are not useful. So I have created width difference and pitch difference variables and relative seats (LAMBDA) are required not no of seats .
Our correlation and t test also suggest that.

Linear equation for Price Premium=
Price Premium = -460.56 + 1.04*Price_economy + 79.94*Flight_duration
-155.22*International + 74.95*diffWidth + 27.84*diffPitch
+985.56*LAMBDA

Residual standard error: 501.4 on 455 degrees of freedom
Multiple R-squared: 0.8509, Adjusted R-squared: 0.8489

After analyzing residual plot and fitted line ,it can be interpreted that this is not best fitting regression line.
So there might be other variables also which are not present in data set ,which can act like a explanatory variables such as headrest, legrest, premium meals, self service bar, more recline, priority check-in etc. And might be playing a role in price of premium economy.

So , now come back to original hypothesis which is factors for difference in premium economy and economy class ticket.
Since Price of Premium has big residuals and some outliers , difference between premium economy and economical class ticket price will also have big residuals and some outliers. And would not be a prefect fit line.

Diff price= price premium - price economy
Variables that would be used in line will also be same because ,we're analyzing this time difference of price , so price premium and price economy variable will not be there as difference automatically is integrating them. And our explanatory variables will include diff pitch ,diff width, premium to total seat ratio,international and flight duration.

Linear equation for difference between premium economy and economy ticket price =

diffPrice = -412.77 + 83.46*Flight_duration - 52.2*International +
85.26* diffWidth + 7.89*diffPitch + 1010.04*LAMBDA

Residual standard error: 501.6 on 456 degrees of freedom
Multiple R-squared: 0.2651, Adjusted R-squared: 0.257

Beta-coefficients whose are not statistically significant (p > 0.05) are

INTERNATIONAL (which is correlated with flight duration, it is generally noticed that mostly premium economy are available in international flights which most probably have big flight duration) and diffPITCH (which is correlated with diffWIDTH, generally if more width is there, there will be more pitch but width will play more important role)

New line excluding those explanatory variables , we fit again
diffPrice = -402.03 + 81.86*Flight_duration + 88.49* diffWidth + 1018.64*LAMBDA

Residual standard error: 500.5 on 458 degrees of freedom
Multiple R-squared: 0.2649, Adjusted R-squared: 0.2601

For eg ; There would be 81.86 USD increment in diff of price for an hour increase in Flight duration.
There would be 88.49 USD increment in diff of price for an Inch increment of difference between width of both class.

Model accounts for 26.5% of variance in diff of price.
Average error in prediction of diff between price will be 500.5

So our interpretation would be, that duration of flight and difference between width of armrests and premium seats to premium+economy seats (Relative seats) will affect the price difference between premium economy and economy class ticket.
There would be other factors too as described above but as of this dataset, these factors will make price diff dependent on them.